

Chapter 5 STAAR Alternate 2



Overview

Participation Requirements

Testing Requirements for Graduation

Test Development

Training

Test Administrations

Testing Accommodations

Scores and Reports

Performance Standards

Scaling

Equating

Reliability

Validity

Measures of Student Progress

Sampling

Test Results

Overview

Due to the COVID-19 pandemic, the testing window for State of Texas Assessments of Academic Readiness (STAAR®) Alternate 2 was extended to allow a local education agency (LEA) to complete testing. The Texas Education Agency (TEA) created the [Coronavirus \(COVID-19\) Support and Guidance](#) webpage to help coordinate the flow of information from the state to districts, help districts solve problems, and provide guidance to aid in districts' decision-making.

The STAAR Alternate 2 is an assessment based on alternate academic achievement standards and is designed for students with the most significant cognitive disabilities. STAAR Alternate 2 was developed to meet federal requirements of both the Elementary and Secondary Education Act (ESEA) and the Individuals with Disabilities Education Act (IDEA). ESEA requires that all students be assessed in specific grades and subjects throughout their academic career, whereas IDEA requires that students with disabilities have access to the same standards as their non-disabled peers and that students be included in statewide assessments.



STAAR Alternate 2 is not a traditional paper or multiple-choice test. Instead, it involves test administrators observing students as they respond to standardized, state-developed assessment items that align to the grade-level [Texas Essential Knowledge and Skills \(TEKS\)](#). Teachers evaluate student performance based on standard scoring instructions embedded into each item on STAAR Alternate 2 and submit student results through the [Texas Assessment Management System](#).

The assessments included in STAAR Alternate 2 are shown in Table 5.1.

Table 5.1. STAAR Alternate 2 Assessments

Grade	Assessed Subject Area/Course				
3	Mathematics	Reading			
4	Mathematics	Reading	Writing		
5	Mathematics	Reading		Science	
6	Mathematics	Reading			
7	Mathematics	Reading	Writing		
8	Mathematics	Reading		Science	Social Studies
9–12	Algebra I	English I English II		Biology	U.S. History

Participation Requirements

Students who receive special education services and have a significant cognitive disability are eligible to participate in STAAR Alternate 2. These students exhibit significant intellectual and adaptive behavior deficits that limit their ability to plan, comprehend, reason, and apply social and practical skills. Such skills include personal care, social problem-solving skills, dressing, eating, using money, and other functional skills across life domains. Students with significant cognitive disabilities require extensive, direct, individualized instruction, as well as a need for substantial supports that are neither temporary nor content-specific. For these students, STAAR Alternate 2 has [specific participation requirements](#) that an admission, review, and dismissal (ARD) committee must carefully consider. Prior to reviewing the eligibility criteria for STAAR Alternate 2, the ARD committee must understand all assessment options, including the characteristics of each assessment and the potential implications of each assessment choice.

When considering STAAR Alternate 2 for a student, the ARD committee must review the five criteria below and indicate whether the description is applicable to the student. For a student to be eligible to participate in STAAR Alternate 2, the answer to all five questions below must be “Yes.” If the answer to any question is “No,” the student is not eligible to participate in STAAR Alternate 2 and must participate in the STAAR

assessments. Each “Yes” answer must be justified by evidence that the student meets the criterion.

1. **Does the student have a significant cognitive disability?** A significant cognitive disability is determined by the ARD committee and must be based on evaluation information performed by a qualified evaluation team. The significant cognitive disability must follow the Texas definition of a significant cognitive disability. A student with a significant cognitive disability has limited potential to reach grade-level expectations, and evidence must be documented in the Participation Requirement information as supported by the student's most recent fully independent evaluation (FIE).
2. **Does the student require specialized, extensive supports to access the grade-level curriculum and environment?** Federal regulations mandate that all students have access to, and are assessed on, grade-level curriculum. To access the state-mandated grade-level or course curriculum, the TEKS, a student with a significant cognitive disability needs specialized academic instruction as well as support throughout the day in independent living skills such as expressing his or her needs, getting from place to place, eating lunch, negotiating social situations, and taking care of personal needs. Specialized supports are not temporary and are required across all environments that the student accesses.
3. **Does the student require intensive, individualized instruction in all instructional settings?** The student needs specialized academic instruction and techniques that are more intense than their other peers with disabilities. A student with a significant cognitive disability not only accesses instruction through non-traditional methods but also classroom assessments.
4. **Does the student access and participate in the grade-level TEKS through prerequisite skills?** Access to the grade-level curriculum is mandated by the federal government. A student with a significant cognitive disability requires access to the TEKS through prerequisite skills that are linked to the grade-level curriculum. These prerequisite skills are listed in the [STAAR Alternate 2 TEKS Curriculum Framework Documents](#). Students eligible for STAAR Alternate 2 may be performing between 3–9 grade levels below their peers.
5. **Is the STAAR Alternate 2 assessment determination based on the student's significant cognitive disability and NOT on any other factors?** The decision to administer STAAR Alternate 2 is NOT based on a student's racial or economic background, English learner (EL) status, excessive or extended absences, location of service delivery, anticipated disruptive behavior or emotional distress, or any other such factors.





Testing Requirements for Graduation

With the passage of the 2009, 81st Texas Legislature, [House Bill \(HB\) 3](#), the relationship between high school courses and participation in the STAAR Alternate 2 end-of-course (EOC) assessments is linked to a student's graduation plan. However, the ARD committee makes educational decisions for a student with a disability, including decisions related to graduation requirements as described in [Texas Administrative Code \(TAC\) §89.1070](#).

Test Development

STAAR Alternate 2 follows the same test-development procedures as other STAAR assessments. However, the test-development process does reflect the unique characteristics of STAAR Alternate 2, specifically its reliance upon scripted items and the learning styles of the STAAR Alternate 2 population.

Assessment Content

Like other STAAR assessments, STAAR Alternate 2 is linked to grade-level TEKS and student expectations for STAAR. The preliminary task in developing the alternate assessment was to link the assessment to curriculum content and expectations. TEA worked with experts in test development, special education, and content to develop curriculum frameworks and vertical alignment documents. The curriculum frameworks list the grade-level TEKS and the associated prerequisite skills for each grade and subject area. The vertical alignment documents link skills and knowledge across grades within the same subject area. After the initial creation of the curriculum frameworks and vertical alignment documents, TEA sought additional input from educator committees and a statewide steering committee that included state assessment experts, parents, advocacy group representatives, related service providers, administrators, and Education Service Center (ESC) professionals.

The next step in developing the STAAR Alternate 2 assessment was to generate [essence statements](#) that summarize the TEKS and student expectations and link the expectations to the prerequisite skills and assessment performance categories. Typically, each grade and subject area assessment contains 10–20 essence statements. From these, specific essence statements are identified for inclusion in the STAAR Alternate 2 assessment each year. The 2020–2021 essence statements were made available to teachers in fall 2020 to allow time for instructional planning and developing standards-based individualized education programs (IEPs) for the school year.

STAAR Alternate Redesign

As a result of the 2013, 83rd Texas Legislature, [HB 5](#), TEA redesigned the STAAR Alternate assessment. To meet requirements of the legislation and to maintain an appropriate assessment for students with significant cognitive disabilities, an item-



based approach to the assessment was implemented for the STAAR Alternate 2. The issues of validity, reliability, fairness, accessibility, and consistency in score interpretations were carefully considered. In addition, the principles of universal design were incorporated in the early stages of test development to develop accessible, non-biased items. Consideration was also given to students' individual response modes, which allow students to show what they know during the assessment in a way that is most consistent during routine instruction.

After prototype items were developed, cognitive labs were conducted to gather information on student performance, engagement, and interaction with the redesigned STAAR Alternate 2 items. Test administrators were interviewed regarding the proposed test design and the feasibility of the assessment for students. The next step was a pilot test to gather further student performance data and a survey of test administrators regarding the STAAR Alternate 2 test items. The data from the cognitive labs, pilot tests, and test administrator surveys were used to develop items for the first operational assessment in spring 2015.

Assessment Item Criteria

In addition to the procedures outlined in [chapter 2, "Building a High-Quality Assessment System,"](#) and described above, nationally-accepted criteria provided guidance during the development of the STAAR Alternate 2 items. Specifically, the following criteria were directly referenced during development of the redesigned STAAR Alternate 2.

- Standard 4.1 of the [Standards for Educational and Psychological Testing \(AERA, APA, NCME, 2014\)](#), which states:

Test specifications should describe the purpose(s) of the test, the definition of the construct or domain measured, the intended examinee population, and interpretations for intended uses. The specifications should include a rationale supporting the interpretations and uses of test results for the intended purpose(s) (p.85).

- Standard 4.8 of the [Standards for Educational and Psychological Testing \(AERA, APA, NCME, 2014\)](#), which states:

The test review process should include empirical analyses and/or the use of expert judges to review items and scoring criteria. When expert judges are used, their qualifications, relevant experiences, and demographic characteristics should be documented, along with the instructions and training in the item review process that the judges receive (p. 88).



- **Universal design.** In incorporating universal design for STAAR Alternate 2, attention was given to (1) students' response modes, allowing students to show what they know and can do, (2) differentiated supports and materials, allowing students to access the content of the assessment, (3) multiple means of engagement to allow students more time to complete the task, (4) meaningful activities, and (5) context (Center for Applied Special Technology, 2018). According to the principles of universal design, each item has precisely defined constructs, has maximum legibility, has maximum readability and comprehensibility, is amenable to accommodations, is accessible and non-biased, and takes into consideration special populations.

Review of Items

During the item-development process for STAAR Alternate 2, educator committees met to complete reviews of every item. The committees were comprised of educators from across Texas, specifically special education experts, special education classroom teachers (including teachers from the Texas School for the Blind and Visually Impaired and the Texas School for the Deaf), teachers of English learners, and general education teachers.

The educator committees focused on the relationship between the grade-level content and the items. Each committee member completed an item judgment form with the following questions regarding each item:

- Does this item measure the reporting category/student expectation/essence statement/prerequisite skill it was designed to measure?
- Is this item an appropriate measure of the TEKS student expectation/essence statement/prerequisite skill?
- Is this item free from bias based on students' personal characteristics such as gender, ethnicity, or disability?
- Would you expect students in your district to have received sufficient instruction by the end of the grade/course to enable them to answer this item correctly?

Feedback from the educator committees was used to revise the STAAR Alternate 2 items as needed.



Training

Resources were provided by TEA, outlining administration procedures, sample items, and online activities prior to the testing window. It was recommended that all personnel who planned to administer STAAR Alternate 2 review these resources prior to the test administration window. In addition, a preview window was offered so that school personnel could review the actual assessment items and apply any accommodations appropriate for their students prior to the test administration window.

STAAR Alternate 2



More information about the STAAR Alternate 2 assessments is available in the [STAAR Alternate 2 Educator Guide](#) on TEA's Student Assessment Division website. This guide is provided to familiarize educators with the STAAR Alternate 2 assessment. It includes test-development information, participation requirements, accommodation information, and sample test questions.

Test Administrations

Students who received remote instruction were allowed to participate in the STAAR Alternate 2 assessment. To ensure equity for all students, a student had to be present at a monitored testing session and be supervised by a trained test administrator. The STAAR Alternate 2 testing window was extended to six weeks to give districts more flexibility in scheduling assessments and more opportunity to test all eligible students while continuing to follow appropriate public health protocols during the administration. For students not able to be in a monitored test session, districts were instructed to mark the student assessment with a score code of "Other."

Administration Procedures

The STAAR Alternate 2 assessment process is designed with scripted test administrator presentation instructions that mirror instructional techniques for a student with a significant cognitive disability. The essence statements, upon which the 2020–2021 STAAR Alternate 2 items were based, were made available in fall 2020 so that test administrators would have general knowledge of the state curriculum students would be held accountable for.

Test administrators take the following steps as part of administering the assessment:

1. Select allowable accommodations that may be appropriate for each student. As part of this step, test administrators determine supports that are needed for individual students during the assessment and accommodate materials as needed.
2. Administer items following the standardized presentation and scoring instructions.



3. Observe and score student performance.
4. Enter scoring information into the Texas Assessment Management System so that each student's assessed performance is recorded.

In rare cases, a student with a severe medical or cognitive impairment may not be able to complete the assessment. For these exceptions, ARD committees determined prior to the administration whether a student's assessment should be coded as a Medical Exception or as a No Authentic Academic Response (NAAR). For both exceptions, the ARD committee makes the determination after reviewing medical and educational records. The decision is documented in the student's IEP, along with evidence to support the determination. A decision not to assess a student is rare. Descriptions of the two categories are provided below. The [STAAR Alternate 2 and TELPAS Alternate Medical Exception Eligibility Requirements](#) can provide more detailed information regarding eligibility.

MEDICAL EXCEPTIONS

Students who are medically fragile and do not attend to or do not tolerate any academic interaction qualify for a medical exception under the following circumstances.

- The student is unable to respond to test questions due to a terminal or degenerative illness.
- The student receives extensive short-term medical treatment due to a medical emergency or serious injury in an accident.
- The student is unable to interact with peers or staff without risk of infection or contamination to himself/herself or others.
- The student is unable to receive sufficient or consistent homebound services due to medical issues.

NO AUTHENTIC ACADEMIC RESPONSE

Students who are unable to produce an authentic academic response due to level of cognition rather than a medical condition may qualify for a NAAR exception under the following circumstances.

- Because of multiple impairments, the student is unable to receive information during instruction and assessment. For example, the student may have a combination of sensory impairments, such as hearing, vision, and or tactile.
- The student is consistently unable to provide an authentic academic response during instruction. His or her behavior may be described by one or more of the following characterizations:
 - The student is unable to demonstrate a meaningful, observable reaction to a specific stimulus.
 - The student exhibits only startle responses.

- The student tracks or fixates on objects at random and not for a purpose.
- The student moves or responds only to internal stimuli.
- The student vocalizes intermittently regardless of changes in environment around him/her.

The [STAAR Alternate 2](#) and [TELPAS Alternate No Authentic Academic Response Eligibility Requirements](#) can provide more detailed information regarding eligibility.

Testing Accommodations

STAAR Alternate 2 is a standardized assessment intended to be appropriate for eligible students in its original intact form. However, ARD committees and test administrators may elect to provide appropriate allowable accommodations to students whose disability precludes them from participating meaningfully in the assessment. Test administrators may use accommodations only if they are routinely provided in classroom instruction and listed in the student's IEP. Some accommodations provided during classroom instruction may not be allowed during testing, as certain accommodations used in the classroom would invalidate the content being assessed or compromise the security and integrity of the test. The accommodations, and guidelines on how they should be applied, appearing in Figure 5.1 below are examples that can be considered for STAAR Alternate 2.



Figure 5.1. STAAR Alternate 2 Allowable Accommodations

Allowable Accommodations
<ul style="list-style-type: none"> ■ Color or highlight images or text ■ Place color overlays on images or text ■ Pair images or text with photographs, picture representations, or real objects of the same content <ul style="list-style-type: none"> • photographs, pictures, or real objects must be as close to the original as possible ■ Attach textured materials to images or text ■ Demonstrate concepts or relationships in images or text ■ Raise or darken the outline in images or text ■ Enlarge images or text <ul style="list-style-type: none"> • magnification devices, photocopying, or computer magnification programs can be used ■ Add braille labels to images or provide text in braille ■ Describe images for students with visual impairments <ul style="list-style-type: none"> • descriptions of images can only include details of what can be seen in the images without comments about the overall impression of the image ■ Provide images or text on separate paper presented one at a time <ul style="list-style-type: none"> • images must be presented in the same order or configuration as they appear in the test booklet ■ Cover or isolate images or text until addressed ■ Use routine picture representations for key words in verbal directions to the student <ul style="list-style-type: none"> • only what is visually presented, stated in text, or supplied in the test administrator instructions can be provided ■ Use calculator, manipulatives, or math tools to arrive at response <ul style="list-style-type: none"> • fraction pieces, geometric shapes, number lines, number charts, money, base-ten blocks, counters ■ Reread sections of the text <ul style="list-style-type: none"> • Follow the guidelines in the “Presentation Instructions” section of the <i>Test Administrator Manual</i> for guidance on repeating presentation instructions and rereading sections of the text. ■ Provide structured reminders <ul style="list-style-type: none"> • personal timers, token systems, color-coded or handwritten reminders, or visual schedules

Scores and Reports

Scoring STAAR Alternate 2 Assessments

STAAR Alternate 2 is scored polytomously using a standard scoring rubric that follows the same process of item administration across all items and is applied to the student performance evaluation information that test administrators submit electronically via the Texas Assessment Management System. Each item is scored according to the level of independence with which a student responds. The scoring rubric is as follows:

- If a student responds correctly to the first presentation of an item, he or she receives a score point of 2. If the student does not respond or responds incorrectly, the item is presented again with allowable teacher assists.
- If the student responds correctly to the second presentation of the item, he or she receives a score point of 1.
- If the student does not respond or responds incorrectly to the second presentation, he or she receives a score point of 0.



Each item is scored in the same manner. Item scores range from 0 to 2. There are 20 scored items per test, resulting in a total test score range of 0 to 40 points.

Description of Scores

Scores for the STAAR Alternate 2 assessments consist of the number of points earned (raw scores), scale scores, and the resulting performance level associated with the student's score.

RAW SCORE

The number of points that a student earns on a STAAR Alternate 2 assessment is the student's raw score. The raw score can be interpreted only in terms of the specific set of test items on that test form. However, because the difficulty of items might vary among test forms over time, raw scores alone cannot be used to compare performance across tests or administrations. To make these comparisons, raw scores must be converted to scale scores.

SCALE SCORE

A scale score is a conversion of the raw score onto a scale that is common to all test forms for that assessment. Scale scores allow for direct comparisons of student performance between specific sets of test items from different test administrations.

The scale score is used to determine whether a student attained Level II: Satisfactory Academic Performance or Level III: Accomplished Academic Performance.

Performance-level cut scores are discussed in the [Performance Standards](#) section of this chapter. Along with raw scores, scale scores for all STAAR Alternate 2 assessments are reported following each test administration.

Scale scores are also used to compare the performance of an individual student with the performance of a demographic group, a program group, an entire campus, or a district at a particular grade. For example, the scores for a Hispanic student can be compared with the average scores of other Hispanic students, all students on campus, or any combination of these aggregations at that grade.

ADDITIONAL PERFORMANCE INFORMATION

Other scores can provide information about a student's relative strengths or weaknesses in core academic areas. For example, reporting category level data can identify areas where a student might be having difficulty. This identification can help campuses plan the most effective instructional interventions.

Report Formats

Two types of reports are provided for the various testing programs, standard and additional. Standard reports are provided automatically to districts, and the information in the standard reports satisfies mandatory reporting requirements.

For more information about scoring and reporting for STAAR Alternate 2, refer to the TEA publication [Interpreting Assessment Reports](#) located on TEA's Performance Reporting website.



Use of Test Results

Reports of STAAR Alternate 2 students are used in

- helping parents monitor the progress their children make;
- informing instructional planning for individual students;
- reporting results to local school boards, school professionals, and the community;
- evaluating programs, resources, and staffing patterns; and
- evaluating district effectiveness in their instructional programs.

Performance Standards

Performance standards relate levels of test performance directly to what students are expected to learn as described in the statewide curriculum.

Performance Levels and Policy Definitions

For the STAAR Alternate 2 assessments, the performance levels are

- Level I: Developing Academic Performance,
- Level II: Satisfactory Academic Performance, and
- Level III: Accomplished Academic Performance.

More detailed descriptions of these performance levels, known as policy definitions, are given below.

LEVEL I: DEVELOPING ACADEMIC PERFORMANCE

Performance in this category indicates that students require additional instructional supports for accessing the curriculum through prerequisite skills. Students acknowledge some concepts, but they demonstrate a minimal or inconsistent understanding of the knowledge and skills that are linked to content measured in this grade or course. Even with continued support, students in this category need significant intervention to show progress in the next grade or course.

LEVEL II: SATISFACTORY ACADEMIC PERFORMANCE

Performance in this category indicates that students are sufficiently prepared for the next grade or course with instructional supports for accessing the curriculum through prerequisite skills. Students demonstrate sufficient understanding of the knowledge and skills that are linked to content measured at this grade or course. Students exhibit



the ability to determine relationships, integrate multiple pieces of information, extend details, identify concepts, and match concepts that are similar. With continued support, students in this category have a reasonable likelihood of showing progress in the next grade or course.

LEVEL III: ACCOMPLISHED ACADEMIC PERFORMANCE

Performance in this category indicates that students are well prepared for the next grade or course with instructional supports for accessing the curriculum through prerequisite skills. Students demonstrate a strong understanding of the knowledge and skills that are linked to content measured at this grade or course. Students exhibit the ability to use higher-level thinking and more complex skills, which includes making inferences, comparisons, and solving multi-step problems. With support, students in this category have a high likelihood of showing progress in the next grade or course through prerequisite skills.

Standard-Setting Process for STAAR Alternate 2

Standards were set for STAAR Alternate 2 in spring 2015. Standard setting for STAAR Alternate 2 involved a process of combining considerations regarding policy, the TEKS content standards, educator knowledge about what students should know and be able to do, and information about how student performance on state assessments aligns with student performance on other assessments. TEA used an evidence-based standard-setting approach (O'Malley, Keng, & Miles, 2012) for the STAAR Alternate 2 program. Using this approach, TEA defined and implemented a nine-step process to establish performance standards for all the STAAR Alternate 2 grades 3–8 and EOC assessments.

Table 5.2 provides high-level descriptions and timelines for the steps in the STAAR Alternate 2 standard-setting process. This nine-step process is modeled after the nine-step STAAR standard-setting process, however some steps happened in a different chronological order than STAAR based on the administration timing and availability of data.



Table 5.2. Overview of the STAAR Alternate 2 Standard-Setting Process

Standard-Setting Step	Description	Timeline
1. Conduct empirical studies.	Analyses of pilot data as well as analysis of score distributions were conducted.	Spring 2015
2. Develop performance labels and policy definitions.	A committee was convened jointly by TEA and the Texas Higher Education Coordinating Board (THECB) to recommend performance categories, performance category labels, and general policy definitions for each performance category. The STAAR Alternate 2 performance labels and policy definitions were adapted from those created by the committee.	January 2015
3. Develop reasonable ranges for performance standards.	The committee considered the policy implications of performance standards, empirical study results, and content recommendations to identify reasonable ranges for performance standards (neighborhoods).	January 2015
4. Develop grade and course PLDs.	TEA and Pearson drafted specific PLDs and educator committees reviewed and edited the PLDs. A goal of the development and review of the specific PLDs was to create an aligned system describing a reasonable progression of skills within each subject area (mathematics, reading, science, and social studies).	January 2015
5. Convene standard-setting committees.	Committees consisting of general education and special education experts with experience in grades 3–12 used performance labels, policy definitions, specific PLDs, and predetermined ranges within which to recommend cut scores for each STAAR Alternate 2 assessment. These committees also provided comments to assist TEA with finalizing the specific PLDs.	April 2015
6. Review performance standards for reasonableness.	TEA reviewed the recommendations across subject areas.	April 2015
7. Approve performance standards.	The Commissioner of Education approved the STAAR Alternate 2 performance standards.	April 2015
8. Implement performance standards.	Once established, performance standards were reported to students for the spring 2015 administration.	May 2015
9. Review performance standards.	Performance standards are reviewed at least once every three years.*	If applicable

*In June 2013, the 83rd Texas Legislature enacted HB 5, which removed the requirement to review performance standards. However, TEA and the Commissioner of Education review statewide performance relative to the standards after each administration.

More details about each of the steps in the STAAR Alternate 2 standard-setting process are provided in the [STAAR Alternate 2 Standard Setting Technical Report](#) available on TEA’s Student Assessment Division website.



Standard-Setting Committees

TEA selected K–12 educators who have had experience with the population of students for whom STAAR Alternate 2 is appropriate and have had content knowledge and classroom experience to serve as standard-setting committee members. The goal of each standard-setting committee was to recommend two cut scores that would define the three performance levels for each of the STAAR Alternate 2 assessments.

In April 2015, educator committees were convened to recommend performance standards for all STAAR Alternate 2 assessments. Committees reviewed STAAR Alternate 2 test booklets, policy definitions, and PLDs. The panelists also received training in the evidence-based standard-setting process that incorporated aspects of the [extended Angoff](#) method, where panelists make judgments about the score needed on each item to demonstrate proficiency (Angoff, 1971; [Hambleton & Plake, 1995](#)).

Committee members were provided reasonable ranges within which performance standards should be set. The ranges were determined using a content review of items, policy definitions, PLDs, and impact data. With this information in mind, committee members were asked to provide recommendations for where the cut scores should be placed to create the three STAAR Alternate 2 performance levels. Three rounds of recommendations were provided, with time for discussion and feedback between rounds. Committee members also participated in an articulation round where they could look at the third-round recommendations across grades/courses and suggest adjustments. TEA used the third-round and articulation recommendations in making final decisions about the performance standards. Performance standards from the April 2015 standard-setting meetings were used to report students' scores in spring 2015.

Outcome of Standard-Setting

The purpose of the standard-setting process is to establish cut scores that reflect the level of performance a student must demonstrate to be classified into a performance level on each STAAR Alternate 2 assessment. These performance standards were approved by the Commissioner of Education in April 2015.

Table 5.3 presents the approved performance standards for STAAR Alternate 2.



Table 5.3. STAAR Alternate 2 Performance Standards

Subject Area	Grade/Course	Level II: Satisfactory	Level III: Accomplished
Mathematics	Grade 3	300	375
	Grade 4	300	387
	Grade 5	300	379
	Grade 6	300	373
	Grade 7	300	375
	Grade 8	300	365
	Algebra I	300	361
Reading/English Language Arts	Grade 3	300	381
	Grade 4	300	384
	Grade 5	300	387
	Grade 6	300	371
	Grade 7	300	371
	Grade 8	300	379
	English I	300	367
	English II	300	366
Writing	Grade 4	300	363
	Grade 7	300	359
Science	Grade 5	300	387
	Grade 8	300	382
	Biology	300	383
Social Studies	Grade 8	300	372
	U.S. History	300	368

Scaling

Scaling is a statistical procedure that places raw scores on a common scoring metric to make test scores comparable across test administrations. As with previous Texas assessment programs, the STAAR Alternate 2 program uses the Rasch Partial-Credit Model (RPCM) to place test items on the same scale across administrations for a given assessment. Once performance standards have been set for an assessment, the Rasch scale is then transformed to a more user-friendly metric to facilitate interpretation of the test scores. Details of the RPCM scaling method used in Texas are provided in [chapter 3, “Standard Technical Processes.”](#)

Reporting Scales

Scale scores for STAAR Alternate 2 assessments are reported on a horizontal scale. Horizontal scale scores allow for direct comparisons of student performance between specific sets of test items from different test administrations for a specific grade and subject.



HORIZONTAL REPORTING SCALES

For STAAR Alternate 2 assessments, a scale score of 300 represents the recommended Level II performance standard. The standard deviation is 60.

It is important to note that although Level II scale score values are fixed across horizontally scaled assessments, Level III scale score values vary across STAAR Alternate 2 assessments. For a given assessment, the Level III scale score value remains constant over time.

The STAAR scale scores represent linear transformations of Rasch proficiency level estimates (θ). Specifically, the transformation is made by first multiplying θ by a slope constant (A) and then adding an intercept constant (B). This operation is described by the equation below:

$$SS_{\theta} = A \times \theta + B \quad (1)$$

where SS_{θ} is the scale score for a Rasch proficiency level estimate (θ). A and B in Equation (1) are referred to as the horizontal scaling constants. These same transformations will be applied each year to the Rasch proficiency level estimates (θ) for that year's set of test items. Values for the horizontal scaling constants are provided in Table 5.4 for the STAAR Alternate 2 grades 3–8 and EOC assessments.



Table 5.4. Horizontal Scaling Constants for STAAR Alternate 2

Subject Area	Grade/Course	A	B
Mathematics	Grade 3	43.9599	297.2305
	Grade 4	42.3406	297.9677
	Grade 5	42.9221	293.4758
	Grade 6	47.3082	293.8972
	Grade 7	45.0653	292.6994
	Grade 8	45.9897	283.5357
	Algebra I	46.1042	287.8285
Reading/English Language Arts	Grade 3	43.5388	283.9777
	Grade 4	45.6246	277.9633
	Grade 5	49.4951	276.0444
	Grade 6	45.0369	277.0312
	Grade 7	45.2817	278.5818
	Grade 8	42.5894	277.6406
	English I	46.1127	288.1951
Writing	Grade 4	49.1207	286.3444
	Grade 7	45.6246	276.9140
Science	Grade 5	43.8943	291.6601
	Grade 8	38.5892	298.4950
	Biology	38.2614	293.1129
Social Studies	Grade 8	41.4662	282.7501
	U.S. History	41.3565	283.7055

Equating

Overview

Used in conjunction with the scaling process, equating is the statistical process that accounts for the slight differences in difficulty across test forms and administrations and allows for the scores from all forms to be placed onto a common scale. By using statistical methods, TEA and the testing vendor equate the results of different tests so that scale scores across test forms and testing administrations can be compared. In the 2014–2015 school year, for the first administration year, TEA conducted one-time live calibrations and base-test reviews of the new STAAR Alternate 2 assessments since the items had not been previously field-tested. Starting in 2015–2016 and going forward, STAAR Alternate 2 equating activities include pre-equating, post-equating, and field-test equating. Refer to [chapter 3, “Standard Technical Processes,”](#) for detailed information about equating.



Pre-Equating

The pre-equating process takes place prior to test administration. It links a newly developed test form onto the scale of the item bank using a set of items that appeared previously on one or more test forms. This permits the difficulty level of the newly developed form to be closely determined even prior to its administration. Thus, the anticipated raw scores that correspond to scale scores at performance standards can be identified. Pre-equating is conducted for all STAAR Alternate 2 tests as part of the test construction process. The pre-equating model is also used in STAAR Alternate 2 when a test form is re-used in a subsequent administration.

Post-Equating

Post-equating uses data from the operational test administration to re-estimate item difficulties and place them onto the scale of the item bank. For the STAAR Alternate 2 assessments, post-equating uses conventional common-item/non-equivalent groups equating procedures.

Field-Test Equating

To replenish the item bank as new tests are created each year, newly developed items must be field-tested and equated to the item bank. Whenever possible, embedded designs are used to field test new items so that test takers are unable to distinguish between the field-test items and operational items on each test form. STAAR Alternate 2 uses this design to embed a cluster of items on every form. This results in student performance data that are more stable.

Reliability

Reliability refers to the expectation that repeated administrations of the same test should generate consistent results. Reliability is a critical technical characteristic of any measurement instrument because unreliable scores cannot be interpreted as valid indicators of students' knowledge and skills.

Reliability for the STAAR Alternate 2 test scores are estimated using statistical measures such as internal consistency, classical standard error of measurement, conditional standard error of measurement, and classification accuracy and consistency. Refer to [chapter 3, "Standard Technical Processes,"](#) for detailed information about reliability.

Internal Consistency

Internal consistency is a measure of the consistency with which students respond to the items within a test. For STAAR Alternate 2, coefficient alpha is used to estimate reliability.



Generally, reliability coefficients ranging from 0.70 to 0.79 are considered adequate, those from 0.80 to 0.89 are considered good, and those at 0.90 or above are considered excellent. However, what is considered appropriate can vary in accordance with how assessment results are used.

Classical Standard Error of Measurement

Classical standard error of measurement (SEM) represents the amount of variance in a score that results from random factors other than what the assessment is intended to measure. The SEM is helpful for quantifying the margin of uncertainty that occurs on every test.

Conditional Standard Error of Measurement

It is important to note that the SEM index provides only an estimate of the average test score error for all students regardless of their individual levels of proficiency. By comparison, conditional standard error of measurement (CSEM) provides a reliability estimate at each score point on a test. More specifically, CSEM is an estimate of the average test score measurement error that is conditional on the proficiency or scale score estimate.

Classification Accuracy and Consistency

Classification accuracy and consistency provide estimates of the accuracy and consistency of student classifications into performance categories based on current test results.

Validity

STAAR Alternate 2 scores are used to make inferences about student achievement. In support of these inferences, evidence is continually collected throughout the development and administration of STAAR Alternate 2 to demonstrate that the assessments measure the intended content. This validity evidence can be categorized as being based on test content, response processes, internal structure, relations to other variables, and the consequences of testing. This validity evidence supports multiple uses of test scores. TEA follows national standards of best practice to continue to build its body of validity evidence for all the STAAR assessments. The Texas Technical Advisory Committee (TTAC) provides ongoing input to TEA about STAAR Alternate 2 validity evidence. The following sections describe the validity evidence that has been collected for STAAR Alternate 2.

Evidence Based on Test Content

Validity evidence based on test content refers to evidence of the relationship between tested content and the construct the assessment is intended to measure. All STAAR assessments, including STAAR Alternate 2, have been designed to align with the



content defined by the TEKS. The STAAR Alternate 2 test-development process plays an integral role in providing validity evidence based on test content for the assessment. The test development process and the evidence collected related to test content support the use of STAAR Alternate 2 scores in making inferences about students' knowledge and understanding of the TEKS.

RELATIONSHIP TO THE STATEWIDE CURRICULUM

At the inception of the STAAR Alternate 2 assessments, a steering committee was convened to review and provide feedback on the alignment of STAAR Alternate 2 assessment tasks to the TEKS. Educator reviews and focus group meetings continue to be a part of ongoing content development with revisions to the STAAR Alternate 2. Both focus groups and educator review meetings have occurred to review and provide feedback on alignment of items and content standards as well as to review and provide feedback on items themselves.

In 2015–2016, an independent third-party analysis of the alignment between items on the 2016 STAAR Alternate 2 tests and the TEKS was conducted to provide information for peer review. This study was intended to provide TEA with information about the degree of alignment between items appearing on the 2016 STAAR Alternate 2 test forms and the Texas curriculum standards. The study concluded that the 2016 STAAR Alternate 2 items demonstrated strong linkage across all grades and content areas. All items were found to have an academic foundation and to have content connections to the grade-level student expectations.

EDUCATOR INPUT

Professional judgments from educator review meetings provided additional content validity evidence. Educators from across the state reviewed the content of every item to validate that each item matched the appropriate content standard. The educator committees included special education experts and general education teachers from each of the twenty educational regions of Texas.

As part of the review meetings, educators considered each item and were asked, “Does this item measure the reporting category, student expectation, essence statement, and prerequisite skills it was designed to measure?” To respond to this question, educators referenced resources such as the TEKS curriculum documents to verify the match of the reporting category, student expectation, essence statement, and prerequisite skills to each item. Across STAAR Alternate 2 items, educator review committees affirmed the relationship between the items and the TEKS. Additional committee input also confirmed that students are provided opportunities to learn the content before the assessment is administered.

Another important source of content validity is evidence related to bias. To be valid, an assessment must not only assess the intended content, but also be free of bias. To provide this validity evidence, educator committees were asked the following question regarding each assessment task: “Is this item free from bias based on students’

personal characteristics such as gender, ethnicity, or disability?” Committee members affirmed that STAAR Alternate 2 items are free from bias.

TEST DEVELOPER INPUT

Item writers and reviewers, who include content experts and special education experts, follow test-development guidelines and item specifications that explain how the content of the assessed TEKS should be measured. At each stage of development, writers and reviewers verify the alignment of the test items with the assessed reporting categories.

Evidence Based on Response Processes

Response processes refer to the cognitive behaviors that are required to respond to a test item. Texas collects evidence to show that the way students respond to items on the STAAR Alternate 2 assessments reflects accurate measurement of the construct.

ITEMS

Texas gathers theoretical and empirical evidence that support the expectation that the way students respond to test items does not add construct-irrelevant variance. Every year, during item reviews, educators evaluate whether the content for a given item is being appropriately assessed and whether students will be able to accurately demonstrate their knowledge of the construct given the items’ planned format. When items are field tested, additional student response data is gathered. Data such as item difficulty, item-total correlations, and item fit are all evaluated. For additional information, see the Item Analyses section of [chapter 3, “Standard Technical Processes.”](#)

SCORING PROCESS

The process used to score items can provide additional validity evidence based on response processes. This type of validity evidence is predicated on accurate scoring. Within the test administrator booklet, test administrators are provided exact scoring rules and scripted instructions for how to present every item to a student. Test administrators are provided resources to prepare for a STAAR Alternate 2 test administration including a period of time directly prior to the testing window in which they can preview the test booklet to prepare for a valid test administration.

Evidence Based on Internal Structure

Texas collects evidence that shows the relationship between items and reporting categories to verify that the elements of an assessment conform to the intended test construct. Texas conducts annual internal consistency studies to gather evidence based on internal structure. The internal consistency of the STAAR Alternate 2 assessments is evaluated every year using coefficient alpha for assessments that have only polytomously scored items. These internal consistency evaluations are made for all students and for student groups such as female, male, African-American, Hispanic, and white students.



Evidence Based on Relationships to Other Variables

Another method Texas uses to provide validity evidence for the STAAR Alternate 2 assessments is analyzing the relationship between performance on STAAR Alternate 2 and performance on other assessments, a process that supports what is referred to as criterion-related validity. Evidence can be collected to show that the empirical relationships are consistent with the expected relationships. STAAR Alternate 2 correlation estimates, which evaluate the strength of the relationship (or the lack of one) between scores on the STAAR Alternate 2 assessments across different content areas (for example, grade 4 mathematics and grade 4 reading, or biology and U.S. history) are calculated.

Evidence Based on Consequences of Testing

Another way of providing validity evidence is by documenting the intended and unintended consequences of administering an assessment. Possible unintended negative consequences can include narrowing of curriculum or instruction to focus on specific learning outcomes that are assessed, or the inappropriate use of test scores by stakeholders. Some of the intended consequences of the STAAR Alternate 2 assessment, based on the requirements in federal and state statutes, are listed below.

- Students with the most severe cognitive disabilities can receive challenging instruction that is linked to state content standards.
- Students with the most severe cognitive disabilities can be included in state assessment programs.
- STAAR Alternate 2 can assess the achievement of students with the most severe cognitive disabilities.

Performance on STAAR Alternate 2 assessments can be used to track the academic progress of students across years.

Measures of Student Progress

Student progress measures provide information beyond performance level by considering performance over time. Whereas performance level information describes students' current achievement, progress measures describe students' achievement in adjacent years. In 2015–2016, the STAAR Alternate 2 progress measure was calculated and reported for the first time.

The STAAR Alternate 2 progress measure provides information about the amount of improvement or growth a student has made from year to year. For STAAR Alternate 2, progress is measured based on a student's stage change from the prior year to the current year. Stage change is determined by: 1) classifying the student's scores from the previous school year and the current school year in terms of the stage of

performance achieved, and then 2) comparing the stages from year to year. Student progress is then categorized as *Did Not Meet*, *Met*, or *Exceeded*.

Because STAAR Alternate 2 testing was canceled for spring 2020 due to the impact of COVID-19, the STAAR Alternate 2 progress measure calculation in 2020–2021 differed slightly from previous years, measuring progress across two years, from 2018–2019 to 2020–2021. The progress measure indicator was also adjusted to give a fairer indication of student progress across a longer period of time. Compared with the historical progress measure indicator, this model includes more *Met* values to account for students making two years of gains.

Steps for calculating a student’s stage change and progress indicator for the STAAR Alternate 2 progress measure can be found in the “[STAAR Alternate 2 Progress Measure Q & A](#)” document on the Progress Measure page of TEA’s Student Assessment Division website.

Sampling

Sampling may be necessary for STAAR Alternate 2 during either test administration or for audits. It can be used for determining a representative sample for analysis or for assignment of test forms. For the test administration, campus assignment of forms uses an annual sampling process wherein a single form is assigned to each campus in such a way that every form has a sample of students responding that are representative of the state demographic makeup. This approach ensures that each campus administers the same form to all students, and that teachers only have to administer a single form.

Test Results

[Appendix C](#) provides scale score distributions and summary statistics, raw score to scale score conversion tables, as well as mean p -values and reliability estimates for all STAAR Alternate 2 assessments administered in spring 2021.

Table 5.5 shows the spring 2021 pass rates for the STAAR Alternate 2 assessments.

Table 5.5. STAAR Alternate 2 Spring 2021 Pass Rates

Subject Area	Grade/Course	Pass Rate
Mathematics	Grade 3	91%
	Grade 4	93%
	Grade 5	93%
	Grade 6	93%
	Grade 7	95%
	Grade 8	95%
	Algebra I	91%
Reading/English Language Arts	Grade 3	84%
	Grade 4	85%
	Grade 5	90%
	Grade 6	89%
	Grade 7	90%
	Grade 8	95%
	English I	94%
	English II	93%
Writing	Grade 4	83%
	Grade 7	91%
Science	Grade 5	95%
	Grade 8	94%
	Biology	96%
Social Studies	Grade 8	94%
	U.S. History	94%

